

Research Ethics - A tiny tutorial

Forum on Digital Ethics in Research, October 18th, 2022

Catherine Tessier

ONERA-DTIS, Université de Toulouse – Member of the French National Pilot Committee for Digital Ethics

catherine.tessier@onera.fr

Deadbots



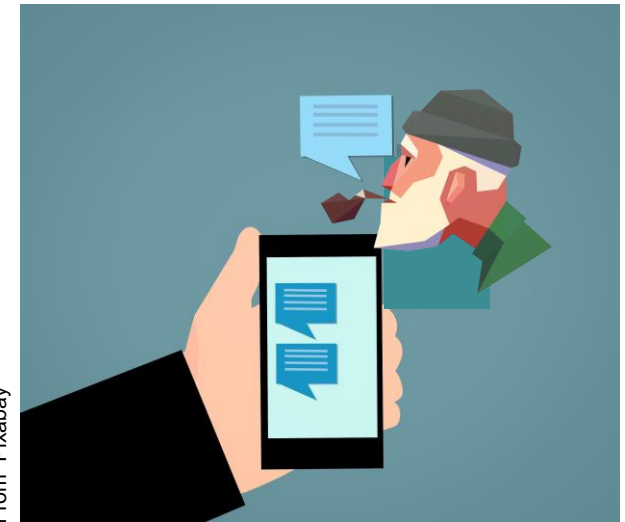
Alice

I enrolled on a new PhD program at Hereafter University. They offered generous grants for students in machine learning to design chatbots replicating the speech of deceased individuals and even generating new phrases that the person has never uttered in their lifetime.



Ben

The Hereafter program looks amazing. I wish I could chat with my grandpa.



From Pixabay

Bee robots



Ben

As for me, the micro-robots are ready to be implemented in the beehive, demo within two months! Our partner Bill Surrogates is looking forward to it.



ROBOROYALE

ROBOTIC REPLICANTS FOR
OPTIMIZING THE YIELD BY
AUGMENTING LIVING ECOSYSTEMS

PROJECT

In RoboRoyale, we aim for a minimally invasive approach to affect large ecosystems in a positive way: We aim to regulate (and also enhance) the egg-laying activity of a honeybee queen by replacing her courtyard by a set of robots that feed and nurse her. By providing an increased flow of proteinaceous food to the queen and guiding her to regions

<https://roboroyale.eu/project.html>

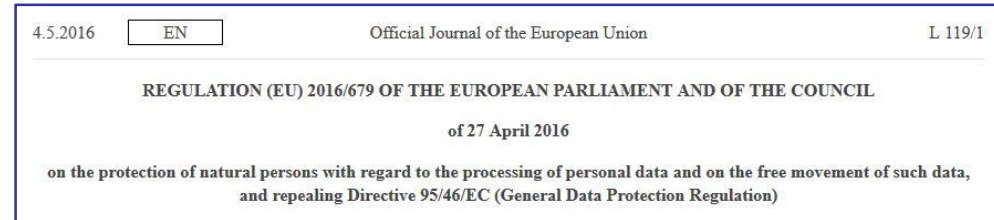
Overview

- Vocabulary
- Thought experiments – Some ethical frameworks
- Back to Alice and Ben
- Some more research ethics questions
- Summary – Some recommendations

Vocabulary

Ethics is not

- the law
 - rules of conduct
 - comes from: sovereign authority
- staff regulations
 - rights and duties
 - come from: organization
- deontology
 - code of conduct
 - comes from: profession



The European
Code of Conduct for
Research Integrity
REVISED EDITION



Morality and values

- **Morality**

- set of **rules** and **principles** according to which one directs one's life and behaviour, considered in relation to good and evil
 - comes from: society (social consensus), religion, yourself
 - normative
- e.g., do not lie, do not steal*



Tumisu - Pixabay

- **Values**

- something desirable, that you stand up for
 - come from: society, yourself
 - a value can be broken
- e.g., honesty, benevolence, respect of human dignity*



Gerd Altmann - Pixabay

- Personal or (more often) collective **thought** about **human behaviours** and **values** they are based on
- **To wonder about** topics, situations, possible decisions and actions **along the way**
- An **approach** that aims at behaving / acting **at best in a specific context**, i.e., at determining what can be considered the right decision or action, according to **arguments** based on **scales of values**
- Ethics does not command, ethics **recommends**



Peggy und Marco Lachmann-Anke - Pixabay

→ Ethical debate is based on **conflicts between values**

[CERNA18] Commission de réflexion sur l'éthique de la recherche en sciences et technologie du numérique d'Allistene - *Proposition de formation doctorale, Initiation à l'éthique de la recherche scientifique*, octobre 2018

Ethics: approaches [CCNE21]

“A collective deliberation, always to be resumed, must make the difference between the possible and the acceptable [...]”.

“New knowledge, by complicating or shifting the scope of our questioning, often increases our uncertainties.”

"Ethical reflection therefore requires us to question the principles that define the very notion of progress [...]"

"Each time [progress] is invoked, we must critically analyse the certainties and convictions that underlie it [...]"

This critical analysis must be based, on the one hand, on values and, on the other, on the most lucid possible assessment of the benefits and risks.”

“But listing these values to define an 'ethical invariant' is not enough. [...] any particular issue requires them to be articulated in order to find the best possible balance when they come into competition.”

[CCNE21] Comité consultatif national d'éthique pour les sciences de la vie et de la santé – *Rapport d'activité 2019-2020*. Août 2021. Translated with DeepL- Emphasis added

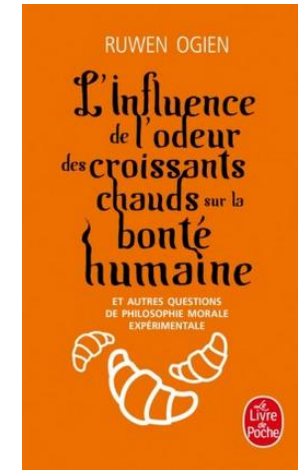
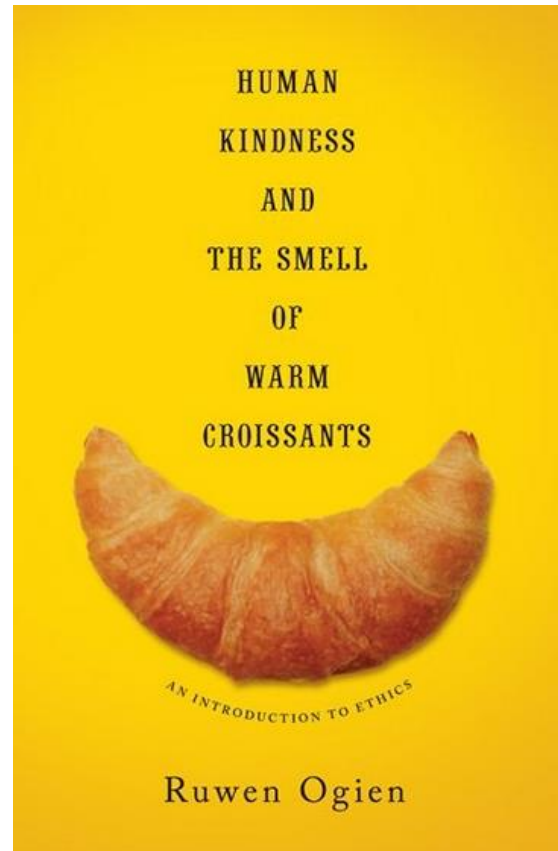
Thought experiments

Some ethical frameworks

Interesting exercises: thought experiments

Moral dilemmas allow us

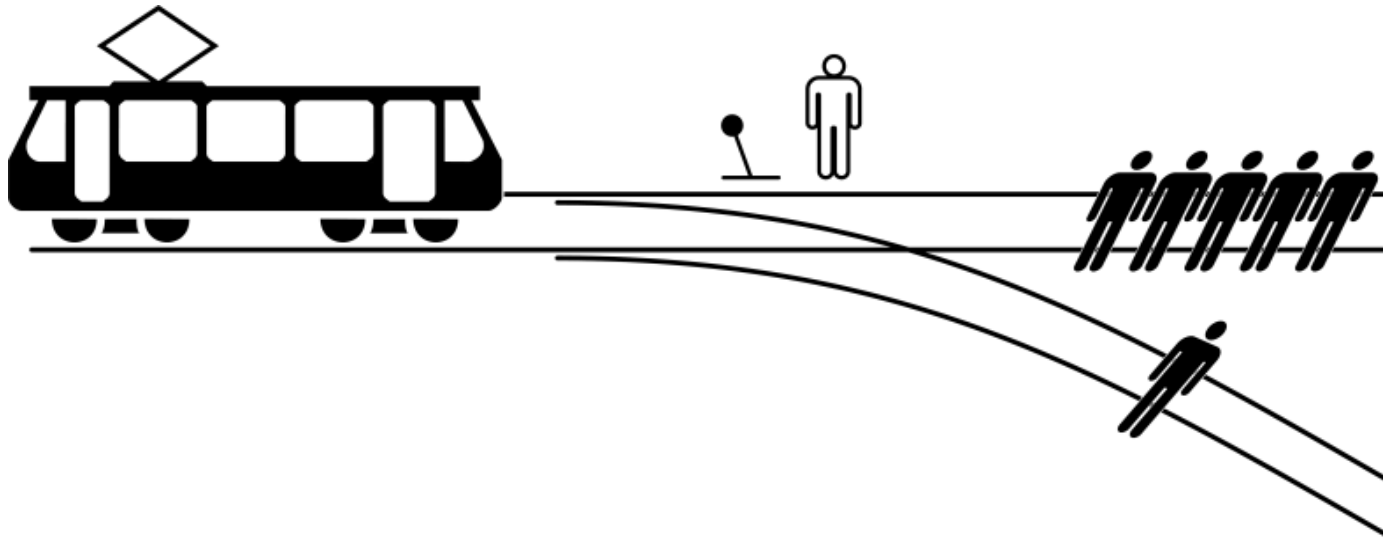
“to identify more clearly the factors that influence our moral judgements”



Thought experiment: the trolley dilemma [Foot02]

What is your decision?

- I pull the lever
- I don't pull the lever



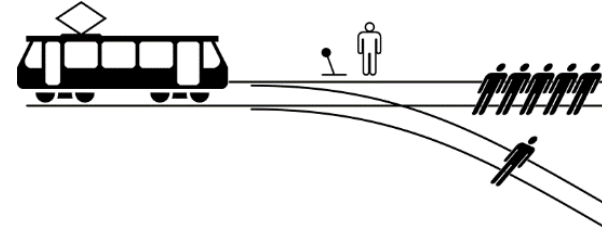
McGeddon, Wikipedia - CC BY-SA 4.0

[Foot02] Foot, Philippa. *Moral Dilemmas and Other Topics in Moral Philosophy*. Clarendon Press. Oxford University Press, 2002.
DOI:10.1093/0199252866.003.0002

Trolley: consequentialism

What is your decision?

- I pull the lever
- I don't pull the lever



Right decision = action whose **outcomes** are “the best”

- positive consequentialism: maximise good (e.g., 5 alive / 1 alive)
- negative consequentialism: minimise evil (e.g., 1 dead / 5 dead)

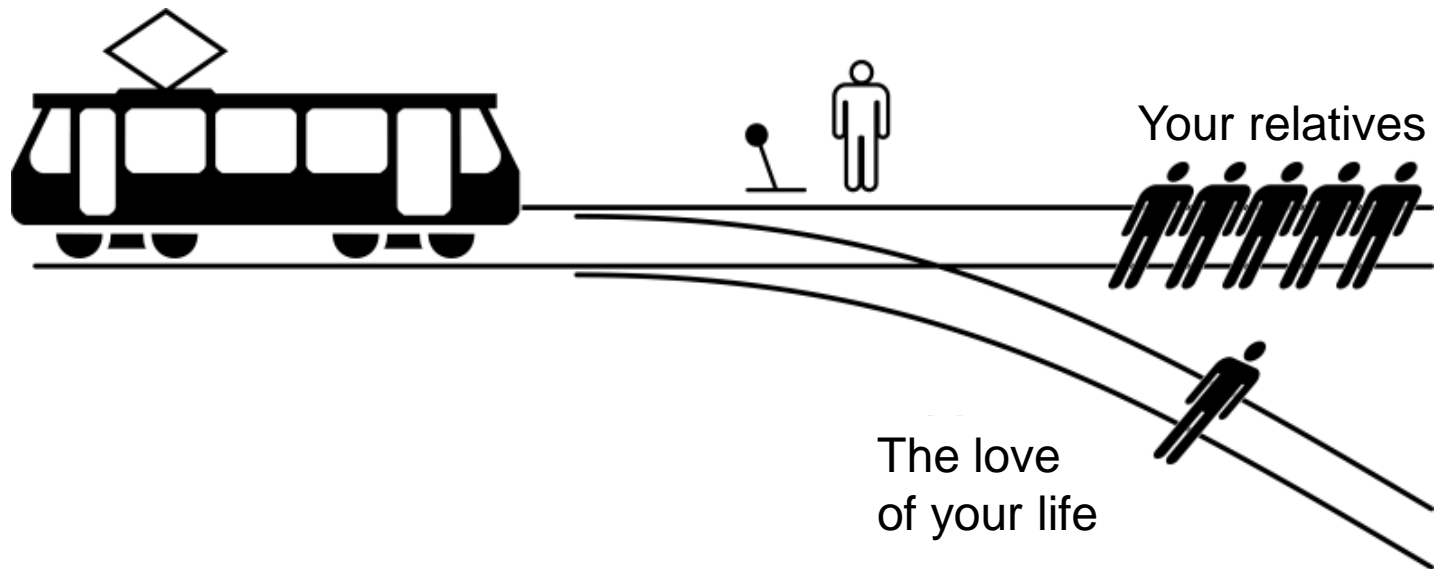
Examples

- Utilitarianism: the greatest good for the greatest number of people
- Selfishness: the greatest good for oneself
- Precautionary principle: risks minimisation

Thought experiment: another trolley dilemma

What is your decision?

- I pull the lever
- I don't pull the lever

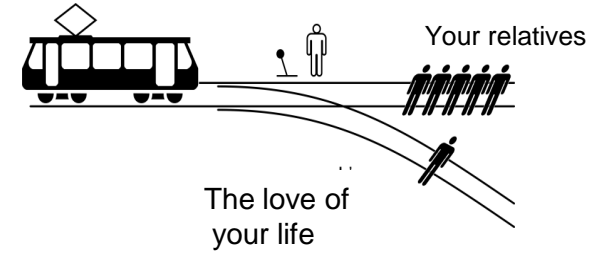


From McGeddon, Wikipedia - CC BY-SA 4.0

Another trolley

What is your decision?

- I pull the lever
- I don't pull the lever



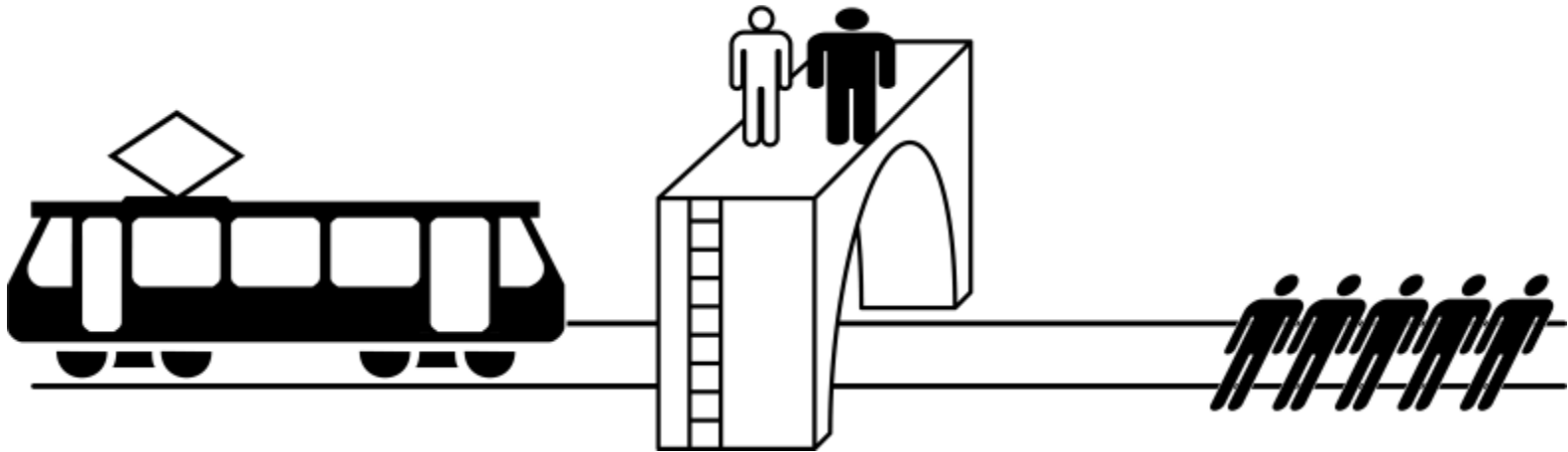
Ethical thought always takes place **within a context**

Which **hierarchy of values** do I consider in this situation?

Thought experiment: the footbridge dilemma [Thomson76]

What is your decision?

- I push the guy
- I don't push the guy



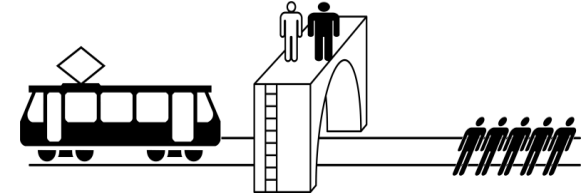
P.Gaborit, from McGeddon, Wikipedia - CC BY-SA 4.0

[Thomson76] Thomson, Judith Jarvis - *Killing, Letting Die, and the Trolley Problem*. *The Monist*, 59(2), 204–217, 1976. DOI:10.5840/monist197659224

Footbridge: deontology

What is your decision?

- I push the guy
- I don't push the guy



Right decision = **action** that respects **some principles** (whatever the consequences)

Kant's categorical imperatives:

“Act only on that maxim whereby you can at the same time will that it should become a universal law”

“Act as to treat humanity whether in your own person or in that of another never as means but always as an ends”

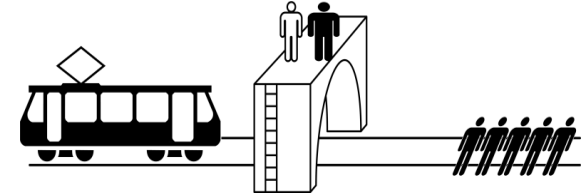
Examples

- Always tell the truth
- Do not make an attempt on somebody's integrity

Footbridge: virtue ethics

What is your decision?

- I push the guy
- I don't push the guy
- I jump



Right decision = action that shows a particular **virtue** of the **agent**

Examples

- honesty, courage, caution, **self-sacrifice**

Back to Alice and Ben

Deadbots



Deontology

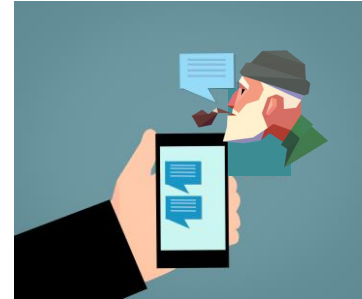
- Respect for human dignity
- Digital twin of a deceased person (process, purpose)
- New outputs that the person never uttered
- Consent
- Business

→ ethical/technical issues

- Data, learning process
- Operating time and conditions
- Vocabulary used
- Name

Consequentialism

- Effects on users
 - mourning process
 - impaired judgement
 - psychological effects (esp. if what is uttered is completely different from what is expected)
- Effects on society



From: Comité national pilote d'éthique du numérique – *Opinion 3: Ethical issues of conversational agents*. Sept. 2021
https://www.ccne-ethique.fr/sites/default/files/2022-05/CNPEN%233-ethical_issues_of_conversational_agents.pdf



PROJECT
In RoboroYale, we aim for a minimally invasive approach to affect large ecosystems in a positive way. We aim to regulate (and also enhance) the egg-laying activity of a honeybee queen by replacing her courted by a set of robots that feed and nurse her. By providing an increased flow of proteinaceous food to the queen and audine her to reuters

Deontology

- Environment well-being
- Sustainability (robots)
- Increase scientific knowledge
- Experiments with (small) animals

→ ethical/technical issues

- Resources necessary to make and operate robots, lifecycle
- Robustness of robots
- Introduction of robots into beehive
- Other uses, misuses

Consequentialism

- Effects on bees
 - behaviours
(local in beehive, global, generalization)
 - survival
- Long-term effects on environment

Some more research ethics questions

Research ethics—some questions 1) research topic

Who suggested the research topic?

→ PhD advisor, funding organisation, boss, me?

Why is this topic worth researching?

→ motivations? e.g., scientific, societal, financial, commercial, strategic

What is my own interest in this topic?

→ my motivations? e.g., I like it, renowned team, lab in a nice city, I will be paid, international opportunities, professional opportunities

Who employs/pays me?

→ what about my independence of mind?

Possible conflict: personal ethics (as a citizen)/professional ethics (as a researcher)

Research ethics—some questions 2) design

- What is intended to be **assigned** to the software/device, i.e., which functions, and why (in the name of what)?
- Is it always possible to express those functions **with mathematical concepts**, and therefore to program them?
 - Is it possible to catch and encode all the **subtleties** of those functions?
 - Is this modelling of reality **scientifically sound**?
- What is the **subjectivity** that is put in the models / that is in the data?

Is it **well located and identified**?

From: M. Gornet, Cl. Kirchner, C. Tessier - Operational fairness for facial authentication systems. *ERCIM News* 131, October 2022

* All face images come from the dataset Labeled Faces in The Wild (LFW)



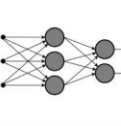
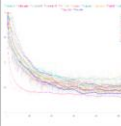
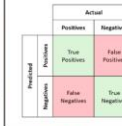
Data Acquisition	Data Processing	Neural Network	Training	Evaluation
				
<ul style="list-style-type: none"> - Image resolution - Number of images - Number of images per person - Labels - Consent - Acquisition technique - Balance - ... 	<ul style="list-style-type: none"> - Normalization - Sampling technique - Augmentation - Number of batches - Train/Test separation - Framing - ... 	<ul style="list-style-type: none"> - CNN - Triplet Loss - Initialization - Layers arrangement - Network depth - ... 	<ul style="list-style-type: none"> - Loss function - Margin - Number of epochs - Hard mining - Regularization - Optimizer - Dropout - Earlystopping - Learning rate and scheduler - Mitigation techniques - ... 	<ul style="list-style-type: none"> - Threshold - Pairs - Metrics - Groups - "Acceptable" value - ...

Figure 1: List of design choices for a facial authentication system and investigated choices (in green)

Research ethics—some questions 3) (future) use

→ What are the **intended uses** of this software/device?

→ Is such a software/device **desirable** for these uses? **In the name of what?**

→ What are the **potential side effects** and **misuses** of this software/device?

Can a researcher anticipate all uses, all effects?

What are their responsibilities?

→ How are the **conflicts between principles or requirements** handled?

e.g.,

Transparency, explicability, predictability / Security

Accuracy / Private life and personal data protection

Decision aid, “autonomy” / Operator’s or user’s autonomy

Off-the-shelf solutions / Sovereignty



Research ethics—some questions 4) users

- Will the users of this software/device **informed/trained**— and how— about
- what it does
 - what it does not do
 - what it does not replace

Will there be clear **cautions** about how to use the device/software?
Can the software/device **explain** to the user what it is doing?



Mohamed Hassan - Pixabay

→ Has the **impact** of the software/device on the user been assessed?
To what extent is it likely to **modify their behaviours**—positively (**intended behaviour**) and negatively (**to the detriment of what**)?

→ **Who controls what is inside the software**, especially code and data (where are they stored)?
Is the user **aware** of that?

Summary – Some recommendations

Summary

Texts about “AI and ethics” suggest that “an ethical AI” is possible

But:

→ Ethics is NOT

- tick-boxes
- compliance
- labels
- certificates
- or law!

→ All the criteria that are set out in the texts cannot be met at the same time!

Ethics is a continuous thought process, along the way
and is a matter of trade-off

Some recommendations (1)

- Avoid thinking only in a utilitarian way, ask also: **in the name of what?**
- Question **motivations**: increase knowledge? help people? develop industry? do as others do? believe a device is better than nothing? etc.
- Question **criteria** such as: be quicker, be cheaper, increase security, less human involved, etc.
- Question the **needs** (are they real needs?)
- Question the **choices** in algorithms (parameters, thresholds, etc.)

Some recommendations (2)



108. Member States should ensure that AI researchers are trained in research ethics and require them to include ethical considerations in their designs, products and publications, especially in the analyses of the datasets they use, how they are annotated, and the quality and scope of the results with possible applications.

Everybody should be trained in ethical thinking

students, researchers, engineers, companies, decision-makers, policy-makers, users...
it is not only a matter for ethics experts or ethics officers!

Ethical thinking should always be associated
with research projects, master theses, doctoral theses, etc.

Every scientific paper should include an ethical discussion

And: ethical issues are tightly interwoven with scientific issues

→ ethical issue can lead to new scientific developments

Who/what can help?

Talk with [colleagues](#) who are aware of ethical issues

Lear about the institution [procedures](#)

Talk with the [ethics officer](#)

Ask the [research ethics committee](#) of the university for advice

Read [papers](#) about ethical considerations in research